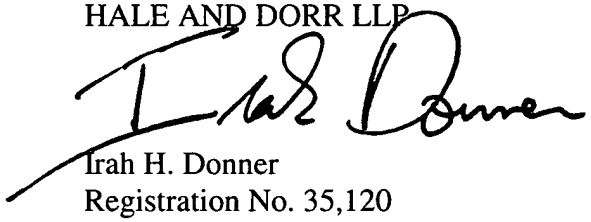
AUTHORIZATION

The Commissioner is hereby authorized to charge any additional fees which may be required for this Amendment, or credit any overpayment to deposit account no. 08-0219.

In the event that an extension of time is required, or which may be required in addition to that requested in a petition for an extension of time, the Commissioner is requested to grant a petition for that extension of time which is required to make this response timely and is hereby authorized to charge any fee for such an extension of time or credit any overpayment for an extension of time to deposit account no. 08-0219.

Respectfully submitted,
HALE AND DORR LLP



Irah H. Donner
Registration No. 35,120

1455 Pennsylvania Avenue, N.W.
Washington, DC 20004-1008
(202) 942-8400 IHD/lrr
Date: 6/4/01

Appendix A

(new claim 21.)

21. A method of calculating similarity or substantial similarity between a first chemical descriptor and at least one other chemical descriptor in a matrix representing a plurality of chemical descriptors, comprising the steps of:

creating at least one chemical descriptor for each compound in a collection of compounds;

preparing a descriptor matrix X, wherein the descriptor matrix comprises the at least one chemical descriptor associated with each respective compound in the collection of compounds;

A1

performing a decomposition of the descriptor matrix to produce resultant matrices used in determining the similarity between the first chemical descriptor and the at least one other chemical descriptor;

determining the similarity between the first chemical descriptor and the at least one other chemical descriptor using at least one of the resultant matrices; and

outputting at least a subset of the at least one other chemical descriptor ranked in order of similarity with respect to the first chemical descriptor.

Appendix B

(complete set of the claims with added claim)

What is claimed is:

1. A method for calculating the similarity of at least one chemical compound to at least one chemical probe, the at least one chemical probe including at least another chemical compound, the method comprising the steps of:

(a) creating at least one chemical descriptor for each compound in a collection of compounds;

(b) representing at least one chemical descriptor for each compound as at least one vector comprising at least one descriptor frequencies;

(c) representing the collection of compound the at least one vector as a first vector of a molecule-descriptor matrix;

(d) performing singular value decomposition of the molecule-descriptor matrix to produce at least one singular matrix;

(e) generating at least one chemical probe descriptor for the at least one chemical probe;

(f) using the at least one singular matrix to transform the at least one chemical probe descriptor of the at least one chemical probe into a first coordinate system at least substantially similar to a second coordinate system of the at least one compound;

(g) calculating the similarity of transformed probes to the compounds in the collection, and

(h) outputting a list of at least a subset of compounds in the collection ranked in order of similarity to the at least one probe.

2. The method as recited in claim 1, wherein said step of creating at least one descriptor includes generating atom pair and topological torsion descriptors from chemical connection tables of the collection of compounds.

3. The method as recited in claim 1, wherein said step of creating at least one descriptor includes creating an index of descriptors and an index of compounds in the collection.

4. The method as recited in claim 1, wherein said molecule-descriptor matrix is denoted as **X**,

wherein said step of performing singular value decomposition includes generating singular matrices as $X = P\Sigma Q^T$ of rank r, and a reduced dimension approximation of *X* defined as $X_k = P_k\Sigma_k Q^T_k$ k<<r, where P and Q are the left and right singular matrices representing correlations among descriptors and compounds respectively, and Σ represents the singular values,

wherein the at least one produced singular matrix includes a pseudo-object denoted as O_F and is calculated from a probe F by $O_F = F^T P_k \Sigma^{-1}_k$, and

wherein said step of calculating the similarity between the pseudo-object O_F and the compounds in collection is computed by taking a dot product of a normalized vector of O_F with each normalized row of P_k.

5. The method as recited to claim 4, wherein said similarity calculating step includes calculating cosine between each pair of vectors.

6. The method as recited in claim 4, wherein said step of performing singular value decomposition includes deriving the reduced dimensional approximation of *X* by setting the *k*+1 through *r* singular values of Σ to zero.

7. The method as recited in claim 4, wherein similarities of the pseudo-object to compounds in the collection is calculated by setting the first k singular values of Σ to one.

8. The method as recited in claim 7, wherein said setting step includes using an identity matrix

9. A method of generating a searchable representation of chemical structures comprising:

(a) generating an index of unique features;

(b) generating a feature-chemical structure matrix including vectors that describe the chemical structures; and

(c) determining correlations between chemical structures based on the generated feature-chemical structure matrix for generating the searchable representation of the chemical structures.

10. The method according to claim 9, wherein the index of unique features include chemical descriptors.

11. The method according to claim 9, further comprising generating the chemical descriptors from connection tables prior to said index-generating step (a).

12. The method according to claim 9, wherein said determining step (c) includes performing singular value decomposition of the feature-chemical structure matrix.

13. The method according to claim 9, wherein the chemical descriptors include at least one of atom pair descriptors, topological torsion descriptors, charge pair descriptors, hydrophobic pair descriptors, inherent atom property descriptors; and geometry descriptors.

14. A computer readable medium including instructions being executable by a computer, the instructions instructing the computer to generate a searchable representation of chemical structures, the instructions comprising:

(a) generating an index of unique features;

(b) generating a feature-chemical structure matrix including vectors that describe the chemical structures; and

(c) determining correlations between chemical structures based on the generated feature-chemical structure matrix for generating the searchable representation of the chemical structures.

15. The computer readable medium according to claim 14, wherein the index of unique features include chemical descriptors.

16. The computer readable medium according to claim 14, further comprising generating the chemical descriptors from connection tables prior to said index-generating step (a).

17. The computer readable medium according to claim 14, wherein said determining step (c) includes performing singular value decomposition of the feature-chemical structure matrix.

18. The computer readable medium according to claim 14, wherein the chemical descriptors include at least one of atom pair descriptors, topological torsion descriptors, charge pair descriptors, hydrophobic pair descriptors, inherent atom property descriptors; and geometry descriptors.

19. The computer readable medium according to claim 16, wherein the instructions further comprise the steps of:

determining whether a user has input a query compound probe;

generating chemical descriptors for the query compound probe;

calculating similarities between the chemical descriptors for the query compound probe and the searchable representation of the chemical structures; and

ranking the chemical structures by similarity to the query compound probe.

20. The computer readable medium according to claim 19, wherein the instructions further comprise the step of:

modifying the query compound probe based on the generated chemical descriptors for the query compound probe.

21. A method of calculating similarity or substantial similarity between a first chemical descriptor and at least one other chemical descriptor in a matrix representing a plurality of chemical descriptors, comprising the steps of:

creating at least one chemical descriptor for each compound in a collection of compounds;

preparing a descriptor matrix X, wherein the descriptor matrix comprises the at least one chemical descriptor associated with each respective compound in the collection of compounds;

performing a decomposition of the descriptor matrix to produce resultant matrices used in determining the similarity between the first chemical descriptor and the at least one other chemical descriptor;

determining the similarity between the first chemical descriptor and the at least one other chemical descriptor using at least one of the resultant matrices; and

outputting at least a subset of the at least one other chemical descriptor ranked in order of similarity with respect to the first chemical descriptor.